心理科学进展 2023, Vol. 31, No. 6, 958–969
Advances in Psychological Science

https://doi.org/10.3724/SP.J.1042.2023.00958

·研究构想(Conceptual Framework)·

混合效应均值–方差模型的建构和样本量规划探索*

刘 玥[1] 方 梵[1] 刘红云[2,3] 雷 怡[1]

([1]四川师范大学脑与心理科学研究院研究院, 成都 610000)
([2]应用实验心理北京市重点实验室; [3]北京师范大学心理学部, 北京 100875)

摘 要 随着研究问题的深入和数据收集手段的进步, 能够合理分析和深入挖掘嵌套结构数据信息的混合效应均值–方差模型(Mixed-Effects Location-Scale Models, MELSM)受到广泛关注。本研究拟通过模拟研究和应用研究, 在贝叶斯框架下探究MELSM的模型建构方法, 并探索MELSM在确定和不确定情境下结合检验力和效应量准确性分析的样本量规划范式, 最终整合上述功能开发简便易用的软件包, 形成MELSM的应用流程, 促进新方法和新技术在心理学研究中的推广应用, 提高研究的生态效度和可重复性, 进而提高研究的整体质量。

关键词 嵌套数据, 混合效应均值–方差模型, 模型建构, 样本量规划

分类号 B841

1 研究背景

在心理学与教育学研究中, 数据常常以层级嵌套的结构呈现。例如, 重复测量设计中试次嵌套于被试, 追踪研究中测量时间点嵌套于个体, 教育学研究中学生嵌套于班级等, 这些多层级结构的数据就是嵌套数据(Nested Data)。嵌套数据对传统的数据分析方法带来了挑战。首先, 嵌套数据由于同一组内的各观测值间不独立, 存在残差非独立性问题, 违背了 *t* 检验、方差分析、回归分析等传统统计方法的假设, 造成结果偏差。其次, 为满足实验控制、因果推断等要求, 研究者需要在不同水平(如试次水平、被试水平等)加入控制变量或预测变量, 这是传统统计方法难以实现的。因此, 越来越多的研究者建议采用线性混合效应模型(Linear Mixed-Effects Models, Hox et al., 2017)实现嵌套数据的分析。然而, 该模型的残差方差同质性假设(以下简称“残差同质性”)在实际应用中常常被违背。例如, Williams 等人(2021)发现, 在探究认知控制的冲突任务中(如, Stroop 任务、Flanker 任务), 被试反应时的变异存在显著的个体差异。忽略残差异质性不仅会造成参数估计偏差, 还不利于研究者深入挖掘有关心理特质稳定性的信息(Williams et al., 2021)。综上, 嵌套数据的分析方法不仅应能考察特质发展变化的趋势及其影响因素(个体间), 还应考察特质在发展中的稳定性及其影响因素(个体内), 为揭示心理现象的本质提供丰富证据。这不仅是传统统计方法无法做到的, 也对线性混合效应模型带来了新的挑战。

为避免残差异质性导致的有偏估计, 并灵活探讨行为特质之间的相互影响、个体间(内)差异及其影响因素, 研究者在线性混合效应模型的基础上拓展提出了一系列广义的模型, 统称为混合效应均值–方差模型(Mixed-Effects Location-Scale Models, MELSM)。MELSM 不需要残差同质性假设, 可以解释不同层级特质的相互作用, 考察特质变异稳定性的影响因素, 从而充分考虑嵌套结构带来的影响, 得到更丰富的研究结果(Williams et al., 2019; Williams et al., 2021)。但是, 研究者在应用 MELSM 时仍存在一些困难。一方面, 由于 MELSM 允许研究者考虑更多随机效应, 忽略必要的随机效应会增大第一类错误率(Barr et al.,

收稿日期: 2022-12-08
* 国家自然科学基金项目(32200920)资助。
通信作者: 刘红云, E-mail: hyliu@bnu.edu.cn;
雷 怡, E-mail: leiyi821@vip.sina.com

2013), 纳入不必要的随机效应又会使模型过于复杂, 导致参数估计困难, 并降低检验力(Judd et al., 2017; Lee, 2018)。因此, 研究者该如何决定应该纳入哪些随机效应从而建构合适的模型? 另一方面, 现有的样本量规划程序(如 G*power, Faul et al., 2007)无法应用于 MELSM, 因此, 研究者应当如何确定适用于 MELSM 的样本量以保证研究结果的可重复性和研究结论的可推广性(Nosek et al., 2022)? 综上, 解决好 MELSM 的模型建构和样本量规划问题, 是促进 MELSM 在心理学研究中推广应用的首要任务。

2　国内外研究现状

2.1　混合效应均值–方差模型的产生及应用

为解决传统线性混合效应模型残差方差异质性的问题, 研究者提出了更广义的 MELSM。该模型可分为两个部分: 均值模型(Location Model, 描述位置的模型, 根据其含义称为均值模型)和尺度模型(Scale Model, 描述尺度的模型, 根据其含义称为方差模型)。

均值模型即为线性混合效应模型的均值部分。以测量(水平 1)嵌套于个体(水平 2)为例, 其一般形式可表示为(Williams et al., 2021):

$$\boldsymbol{Y_i} = \boldsymbol{X_i \beta} + \boldsymbol{Z_i b_i} + \boldsymbol{e_i}, \quad (1)$$

其中, $\boldsymbol{Y_i}$ 是 $n_i \times 1$ 的列向量, 表示水平 2 中个体 i 的结果变量, 其中 n_i 表示个体 i 水平 1 的测量次数, $\boldsymbol{X_i}$ 是 $n_i \times p$ 的矩阵, 第一列是 1, 表示个体 i 的截距, 第 2–p 列是 p–1 个预测变量, $\boldsymbol{\beta}$ 是 $p \times 1$ 的列向量, 表示截距和 p–1 个预测变量的固定效应。$\boldsymbol{Z_i}$ 是 $n_i \times q$ 的矩阵, 第一列是 1, 表示个体 i 的截距, 第 2–q 列是 q–1 个具有随机效应的水平 1 预测变量, $\boldsymbol{b_i}$ 是 $q \times 1$ 的列向量, 表示截距和 q–1 个预测变量的随机效应。残差 $\boldsymbol{e_i}$ 是 $n_i \times 1$ 的列向量, $\boldsymbol{e_i} \sim N_{n_i}(0, \boldsymbol{R_i})$, $\boldsymbol{R_i}$ 是 $n_i \times n_i$ 的协方差矩阵。$\boldsymbol{R_i} = \sigma_e^2 \boldsymbol{I}_{n_i}$, 其中 $\boldsymbol{I}_{n_i}$ 是 $n_i \times n_i$ 的单位矩阵。线性混合效应模型通常假设 $\boldsymbol{R_i}$ 满足同质性假设, 即残差在给定随机效应的情况下条件独立。然而, 很多研究发现这一假设在实际中常无法满足(Hedeker et al., 2008)。因此, MELSM 放宽了残差同质性的限制, 在方差模型中允许水平 1 残差方差异质。

方差模型的定义如下

$$\boldsymbol{\sigma}_{e_i}^2 = exp(\boldsymbol{W_i \tau} + \boldsymbol{A_i t_i}), \quad (2)$$

其中, 个体 i 的残差方差 $\boldsymbol{\sigma}_{e_i}^2$ 是 $n_i \times 1$ 的列向量, $\boldsymbol{W_i}$ 是 $n_i \times s$ 矩阵, 第一列是 1, 表示个体 i 的截距, 第 2–s 列是 s–1 个预测变量, $\boldsymbol{\tau}$ 是 $s \times 1$ 的列向量, 表示截距和 s–1 个预测变量的固定效应。$\boldsymbol{A_i}$ 是 $n_i \times a$ 的矩阵, 第一列是 1, 表示个体 i 的截距, 第 2–a 列是 a–1 个具有随机效应的水平 1 预测变量。$\boldsymbol{t_i}$ 是 $a \times 1$ 的列向量, 表示截距和 a–1 个预测变量的随机效应。方差模型能够有效避免残差异质性对参数估计造成的偏差, 并且解释异质的原因。例如, 将这种模型应用于试次嵌套于被试的实验研究, 可以探索个体内反应变异性更大(或更小)的原因。又例如, 可在此基础上对均值模型水平 2 随机部分的方差也加入预测变量, 探索水平 2(个体间)差异的影响因素(Blozis et al., 2020)。将这种模型应用于学生嵌套于学校的发展与教育研究中, 可以同时探索影响学校内、学校间成绩变异的因素, 帮助改进教学实现教育公平(Williams et al., 2022)。

在 MELSM 的整体框架下, 研究者还可以计算均值模型和方差模型随机效应的相关, 进一步丰富研究结果。例如, 在试次嵌套于被试的实验研究中, 均值模型斜率随机部分与方差模型截距随机部分的相关, 描述了实验效应更强的被试, 倾向于有更一致(或不一致)的反应。此外, 也可以对随机部分的协方差加入预测变量进行解释。

MELSM 具有很强的可拓展性。很多研究者针对不同的研究问题, 基于 MELSM 拓展出丰富的形式。例如, 在均值模型和方差模型中加入非线性部分, 以反映不同个体的学习轨迹变化并探索其影响因素(Williams et al., 2019)。此外, 目前还拓展出适用于结果变量为顺序变量(Hedeker et al., 2016), 时间–事件截断数据变量(Courvoisier et al., 2019), 半连续变量(例如有较多 0 的数据, Blozis et al., 2020)的形式, 以及数据整体结构为动态数据(Rast & Ferrer, 2018), 交叉分类数据(Brunton-Smith et al., 2017), 三水平嵌套设计数据(Lin et al., 2018)的形式。总的来说, MELSM 的优势在于能够同时考察因变量发展变化, 及其变异一致性程度的影响因素, 使其在心理学实验研究、追踪研究等领域的应用越来越广泛。

2.2　混合效应均值–方差模型建构的研究现状

MELSM 的均值模型和方差模型都可能包含随机效应, 如何选择合适的随机效应建构模型, 是模型应用面临的首要问题。

错误定义模型可能会导致错误的推论。一方面，忽略必要的随机效应会导致有偏差的结果。Barr 等人(2013)推荐在没有理论假设的情况下，使用包含尽可能多随机效应的模型。González 等人(2014)支持了他们的观点，发现忽略必要的随机效应将使得残差独立性假设，正态性假设和方差齐性假设无法满足，最终导致标准误的错误估计。另一方面，定义错误的随机效应也会带来严重后果。Lee (2018)认为应当正确定义线性混合效应模型的随机部分。如果将固定效应定义为随机效应，会增加参数估计的误差，导致估计的方差为负(Baird & Maxwell, 2016)，降低模型检验力(Matuschek et al., 2017)。如果随机部分太多，会导致模型不易收敛(Judd et al., 2012)。因此，忽略存在的随机效应和增加不必要的随机效应都可能对参数估计带来不利影响，应当通过模型选择确定合适的随机效应，建构正确的模型(Brauer & Curtin, 2018; Martínez-Huertas et al., 2021)。

模型建构首先应考虑研究设计对随机效应的理论假设。如果缺乏足够的理论依据，则考虑通过数据驱动的方式选择合适的模型。目前，已有关于 MELSM 的研究几乎都是直接定义模型，即使进行了模型选择，也仅关心方差模型是否存在残差异质性(Williams et al., 2021)，或者模型中非线性部分及其相应的随机效应是否成立(Williams et al., 2019)，并没有在典型、完整 MELSM 的框架下探讨模型选择的方法。

此外，探讨模型选择和建构的方法需要结合适用于 MELSM 的参数估计方法。在极大似然估计框架下，可采用似然比检验(Likelihood Ratio Test, LRT)，基于信息准则(例如 AIC, BIC)等方式实现线性混合效应模型的比较和选择(Lee, 2018)。然而，MELSM 的复杂程度较高，极大似然估计会导致参数估计不易收敛。已有关于 MELSM 的研究大多都应用贝叶斯方法实现参数估计(如 Rast & Ferrer, 2018; Williams et al., 2020)。贝叶斯估计可以灵活实现复杂模型的参数估计。在贝叶斯估计框架下，很多常用的模型选择方法已不再适用，需要对传统的拟合指标进行拓展。贝叶斯框架下的拟合指标可分为两类。一类是直接在贝叶斯估计框架下定义的拟合指标。例如，DIC 指标(Deviance Information Criterion, Spiegelhalter et al., 2002)利用了参数后验分布信息计算模型与数据的拟合程度，并包含了惩罚模型复杂度的因子；贝叶斯因子改进了频率学派中使用 p 值的弊端，反映了当前数据将先验概率更新为后验概率过程中的变化(Hoijtink et al., 2019)；后验预测 p 值(Posterior Predictive p-value, PPP, Gelman et al., 1996)反映了在所有马尔科夫链蒙特卡洛(Markov Chain Monte Carlo, MCMC)迭代中，后验预测差异统计量大于当前数据差异统计量的比例。另一类指标是将评价结构方程模型拟合程度的近似拟合指标拓展到贝叶斯框架下。近似拟合指标避免了严格检验的缺陷，能够容忍很小程度的错误。Asparouhov 和 Muthén (2021)提出了将 CFI, TLI, RMSEA 拓展到贝叶斯框架下的方法，其优势在于能够得到指标的可信区间(Credible Interval)，进而应用可信区间而非点估计值进行模型比较。目前，尚没有研究考察不同贝叶斯拟合指标在 MELSM 模型选择中的表现。本研究拟比较 DIC, PPP, 贝叶斯因子，CFI, TLI, RMSEA 对 MELSM 模型选择的结果。

2.3 样本量规划的研究现状

小样本量导致检验力不足的情况广泛存在于各类学科领域的实验研究中(Brysbaert & Stevens, 2018)。低检验力会造成基于 p 值的结果可重复性低(胡传鹏 等, 2016)。已有研究大多针对线性混合效应模型开展检验力分析以规划样本量。仅有 Walters 等人(2018)关注了 MELSM 的检验力。但是，他们在均值模型没有预测变量的情况下只关注了方差模型识别残差异质性或预测变量的检验力，方差模型的预测变量也没有随机斜率，并没有在完整 MELSM 的框架下考察模型固定效应的检验力以实现样本量规划。

与此同时，美国统计协会强调应避免仅报告显著性，而加入对参数估计准确性(Accuracy in Parameter Estimation, AIPE，主要指效应量参数，以下简称“效应量准确性”，Halsey et al., 2015; Maxwell, 2004)的考察(Wasserstein & Lazar, 2016; 温忠麟 等, 2016)。综上，样本量规划应不仅应满足检验力的要求，还应基于效应量准确性分析展开。效应量准确性分析的核心是控制效应量置信区间宽度，越窄表明其估计越准确。然而，目前尚没有研究针对 MELSM 同时结合检验力分析和效应量准确性分析结果规划样本量。

在以上传统样本量规划研究中，往往面临来

自三个方面的不确定性问题(Pek & Park, 2019, 2022)。(1)总体效应量的不确定性。例如, 在实际中为了实现基于回归模型的样本量规划, 研究者往往使用预研究或者前人研究得到的回归系数点估计值, 来代替回归系数的真值(总体效应量)。然而, 基于不同的样本(研究)会得到不同的回归系数, 造成总体效应量的不确定性。(2)样本变异造成的不确定性。即不同研究中具体使用的样本是存在差异的, 而样本量规划没有考虑具体样本的特征, 仅能够给出一般的样本量建议。(3)模型选择造成的不确定性。当存在多个备选模型时(如MELSM 的建构过程), 为了实现样本量规划, 研究者往往预设选择的即为正确模型。然而, 数据分析时选择的模型可能并不能正确代表实际的数据结构, 带来了模型选择造成的不确定性。

常用的基于蒙特卡洛模拟进行检验力分析的范式仅能够通过重复抽样的方式处理样本变异造成的不确定性, 忽略了总体效应量的不确定性和模型选择造成的不确定性, 会导致不准确的结果。例如, Liu 和 Wang (2019)证明, 在未考虑不确定性的情况下进行样本量规划进而开展的研究会面临检验力不足的后果。因此, 在检验力和效应量准确性分析中考虑不确定性问题, 可以更好地代表实践中研究设计面临的现实困境, 也能保证样本量规划结果更加准确可靠, 这使得越来越多的方法研究者关注了不确定情况下的样本量规划问题(如 Anderson et al., 2017; Liu & Wang, 2019)。Pek 和 Park (2019, 2022)提出了贝叶斯经典混合方法(Bayesian-classical Hybrid Approach)并开发了相应软件包, 为解决不确定性问题提供了可行路径。但是, 其研究没有针对 MELSM, 只考虑了检验力分析而未考察效应量准确性分析, 并且在解决模型选择造成的不确定性时, 采用的模型平均的方式在实际应用中不如模型选择广泛(如 Barr et al., 2013; Lee, 2018)。

3　问题提出

结合 MELSM 的理论和实证研究现状可知, 目前有关 MELSM 的模型建构和样本量规划问题仍未能得到充分解决。具体体现在以下几个方面:

首先, 已有研究大多是直接定义模型, 没有在完整的 MELSM 框架下讨论模型建构方法。与线性混合效应模型相比, MELSM 可以加入方差模型的随机效应, 备选模型的增加为模型选择带来了更大困难(Williams et al., 2019), 在实际研究中研究者往往需要通过模型选择来确定各随机效应的取舍, 那么, 对于完整 MELSM, 研究者应该按照何种顺序进行模型建构? 另一方面, 以往的研究者大多基于极大似然估计框架下的拟合指标探讨线性混合效应模型的建构, 然而复杂随机效应模型导致的不收敛的问题促使研究者考虑更合适的参数估计方法。在更适用于 MELSM 的贝叶斯估计框架下, 各拟合指标表现如何? 对上述问题的研究将有效解决 MELSM 应用中的模型建构问题。

其次, 已有研究在样本量规划方面存在许多不足。第一, 已有研究针对 MELSM 开展检验力分析时仅基于简单模型考察方差模型中识别残差异质性或方差模型中预测变量的检验力, 导致研究结果难以推广到完整 MELSM 框架中, 且无法同时考察均值模型和方差模型的固定效应的检验力。第二, 以往的样本量规划主要都基于检验力分析展开, 鲜有研究提出同时考虑检验力和效应量准确性分析实现样本量规划的范式, 更没有研究探索在不确定性情境下, 结合以上二者科学规划样本量的范式, 导致实际中研究的可重复性低, 检验力和效应量准确性达不到预期水平。此外, Pek 和 Park (2019, 2022)提出的贝叶斯经典混合方法采用模型平均的方法不如模型选择应用广泛, 由此得到的样本量建议实用性不强。因此, 在检验力分析的基础上, 如何基于效应量准确性分析开展 MELSM 的样本量规划? 进一步, 在考虑不确定性问题的情况下, 如何开展基于 MELSM 的样本量规划? 如何借助更常用的模型选择方式, 解决模型选择造成的不确定问题? 上述问题的研究有助于完善样本量规划的理论研究, 帮助研究者得到更可靠的样本量建议。

最后, 以往开发的应用于混合效应模型的软件包功能单一, 只能实现模型参数估计及建构、检验力分析、效应量准确性分析中的某一项功能, 尚没有软件包能灵活实现 MELSM 的参数估计及模型选择, 在未考虑和考虑了不确定性问题的情境下同时进行检验力和效应量准确性分析以规划样本量等系列功能。为促进 MELSM 的广泛应用, 研究者需要功能完善、使用便利的应用程序, 用于实现样本量规划和模型建构等功能。

4 研究构想

本研究基于 MELSM 的应用需要，围绕 MELSM 的模型选择与建构、样本量规划方法展开系统研究。整体上采用理论研究和应用研究相结合的方法，具体研究流程如图 1 所示。

研究 1、研究 2 和研究 3 主要采用理论推导和模拟研究的方式。参考 Williams 等人(2021)的研究，模拟研究考察均值模型和方差模型最多各包含 1 个自变量，且为相同自变量的情况(开发的软件包将容纳更多常用的模型)。按照模型的复杂程度，主要考虑 6 种嵌套模型。表 1 展示了各模型的主要特征和模型间的嵌套关系，为简化研究，暂不考虑均值模型和方差模型随机效应的相关(Arend & Schäfer, 2019)。其中，模型 1 不包含随机效应，模型 2 在模型 1 的基础上增加均值模型的随机截距，模型 3 在模型 2 的基础上增加均值模型的随机斜率，模型 4 在模型 3 的基础上增加方差模型的随机截距，模型 5 在模型 4 的基础上增加方差模型预测变量的固定斜率，模型 6 在模型 5 的基础上增加方差模型预测变量的随机斜率。因此，模型 1~3 均假设残差同质性假设成立，模型 4~6 则假设残差同质性假设不同程度的被违背。

研究 4 基于心理学实验研究和调查研究等实证研究探索样本量规划和模型建构方法的应用。该研究分别开展了基于 Stroop 研究范式的心理学实验研究，以及探索数学学习自我效能感和教师认知激发教学策略使用程度对数学成绩影响的教育心理学调查研究。

本研究采用的统计软件主要包括用于贝叶斯估计的 R 软件包 brms(Bürkner, 2017)，以及自行开发的软件包。各研究的具体方案如下。

4.1 研究 1：混合效应均值−方差模型的选择与建构研究

研究 1 将基于适用于复杂模型参数估计的贝叶斯方法及相应的拟合指标，探索实现 MELSM

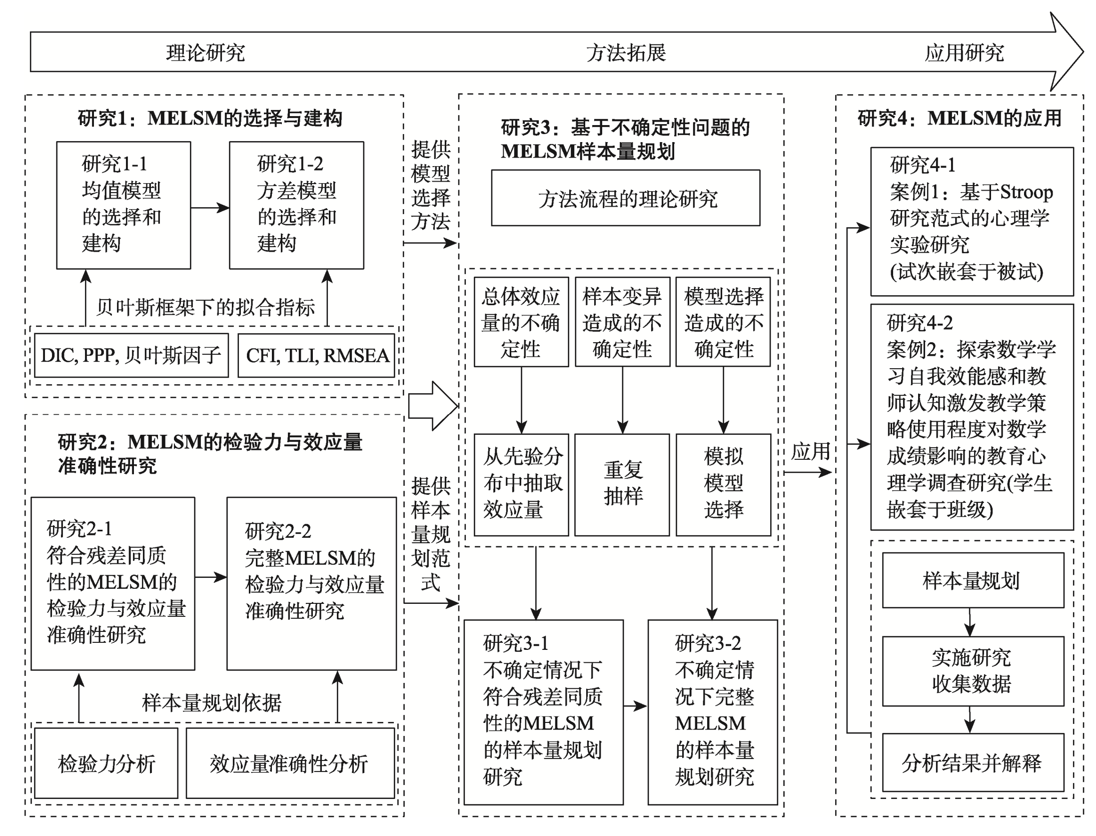


图 1 研究流程示意图

表 1　模拟研究中主要考虑的 6 种模型主要特征

模型	均值模型				方差模型			
	固定效应		随机效应		固定效应		随机效应	
	截距	斜率	截距	斜率	截距	斜率	截距	斜率
均值模型的选择与建构								
模型 1	√	√			√			
模型 2	√	√	√		√			
模型 3	√	√	√	√	√			
方差模型的选择与建构								
模型 4	√	√	√	√	√		√	
模型 5	√	√	√	√	√	√	√	
模型 6	√	√	√	√	√	√	√	√

模型建构的基本途径。主要思路是基于包含不同随机效应的真实模型(模型 1~6)产生数据，然后使用贝叶斯框架下的拟合指标比较各备选模型，确定数据支持的模型。最后，评价所有重复模拟中选出的模型与真实模型的一致性程度，总结各拟合指标的优缺点和适用范围，从而选出稳健性较好的拟合指标用于研究 3。由于在实际中同时考虑均值模型和方差模型会涉及较多的备选模型，因此，为简化流程，本研究拟采用先选择确定最合适的均值模型，再选择确定最合适的方差模型的思路，并考察这种思路的可行性。研究 1 分为两个子研究：

研究 1-1 包括两个情境，分别考察数据符合和不符合残差同质性假设时，均值模型的比较和选择结果。模拟条件包括水平 1 样本量(10, 30, 70, 100, 300)，水平 2 样本量(20, 50, 300, 800)，均值模型自变量的效应量(0.2, 0.5, 0.8)和均值模型随机斜率的方差(0.01, 0.09, 0.25)，共形成 5×4×3×3 = 180 种模拟实验条件水平的组合(即处理)。水平 1 样本量中，n = 10 的水平代表了 Lee (2018)的研究中使用 Laplace 接近方法没有收敛问题的条件，n = 300 的水平代表了 Schultzberg 和 Muthén (2018)关于动态结构方程模型样本量规划研究中测试时间点设置的最大水平；水平 2 样本量中，N = 20 的水平接近 Lee (2018)总结的类似实验设计所使用的被试量最小值(16)，N = 800 的水平接近 Lee (2018)模拟研究中设置的 1000 名被试的水平，目的是为了探索大样本条件对效应量估计准确性和检验力提高的作用。最小到最大样本量水平之间的变化参考了同类样本量规划研究(如 Schultzberg & Muthén, 2018)。均值模型自变量的效应量参考了 Cohen 的 d 值小、中、大水平(Barr et al., 2013; Lee, 2018)。均值模型随机斜率的方差参考了 Arend 和 Schäfer (2019)研究中设置的水平。对于模拟设计的每种组合，参考大多数检验力分析的研究，基于各产生数据的模型，重复生成数据 1000 次(如 Thoemmes et al., 2010; Zhang, 2014)。情境 1 基于模型 1~3 产生数据，情境 2 基于方差模型为模型 4 的方差模型，均值模型分别为模型 1~3 的均值模型，这 3 个模型产生数据。两个情境中拟合的备选模型均为模型 1~3。应用贝叶斯估计方法估计参数。结合敏感性分析结果确定合适的参数先验分布设置。参考前人类似研究(Depaoli & Clifton, 2015; van Erp & Browne, 2021)，拟对于方差比较两种先验分布：无信息先验(逆 Gamma 分布)，稳健分布先验(混合逆 Gamma 分布)。回归系数部分拟参考类似研究，采用标准正态分布(Depaoli & Clifton, 2015; van Erp & Browne, 2021)。模型拟合后，基于拟合指标选择最佳模型，在每种产生模型下计算各指标正确选择模型的比例，选出稳健性较好的拟合指标。研究比较的拟合指标包括 DIC, PPP, 贝叶斯因子, CFI, TLI, RMSEA。

研究 1-2 考察在均值模型已经正确建构的情况下，当残差同质性假设不成立时，方差模型的比较和选择结果。模拟条件与研究 1-1 相同，共形成 180 种模拟实验条件水平的组合(即处理)。对于模拟设计的每种组合，分别基于模型 4~6，重复生

成数据 1000 次。拟合的备选模型为模型 3~6, 研究比较的拟合指标和分析流程均与研究 1-1 相同。

4.2 研究 2：混合效应均值−方差模型的检验力与效应量准确性研究

研究 2 将基于检验力和效应量准确性分析实现样本量规划, 使得样本量同时满足二者要求。主要思路是采用与产生模型相同的拟合模型, 在不同样本量条件下, 基于蒙特卡洛模拟方式计算检验力, 并运用基于后验分布的方法计算效应量的 95%可信区间。研究 2 分为两个子研究, 分别以模型 3 (符合残差同质性的 MELSM, 即线性混合效应模型)和模型 6 (完整 MELSM)为产生数据和拟合数据的模型, 检验力和效应量准确性的分析对象分别为均值模型中预测变量的固定效应, 以及均值模型和方差模型中预测变量的固定效应。

研究 2-1 考察符合残差同质性的 MELSM 的检验力与效应量准确性。模拟条件包括水平 1 样本量(10, 25, 50, 75, 100, 150, 200, 300), 水平 2 样本量(20, 30, 50, 75, 100, 150, 200, 300, 800)和均值模型自变量的效应量(0.2, 0.5, 0.8), 共形成 8×9×3 = 216 种模拟实验条件水平的组合(即处理)。对于每种模拟实验条件水平的组合, 基于模型 3, 重复生成数据 10000 次。重复次数的设置参考了检验力分析相关研究中敏感性分析的结果(Pek & Park, 2019, 2022)。拟合模型 3, 针对均值模型中预测变量的固定效应, 计算检验力、效应量估计值的 95%可信区间宽度、95%可信区间对真值的覆盖率。最后得到同时满足检验力高于 0.8, 95%可信区间宽度较窄, 95%可信区间对真值的覆盖率在 92.5%到 97.5%之间的样本量作为推荐样本量。

研究 2-2 考察完整 MELSM 的检验力与效应量准确性。模拟条件在研究 2-1 的基础上, 增加方差模型自变量的效应量(0.2, 0.5, 0.8), 共形成 8×9×3×3 = 648 种模拟实验条件水平的组合(即处理)。对于每种模拟实验条件水平的组合, 基于模型 6, 重复生成数据 10000 次。拟合模型 6, 针对均值模型和方差模型中预测变量的固定效应进行分析, 分析流程和结果评价与研究 2-1 相同。

4.3 研究 3：基于不确定性问题的混合效应均值−方差模型样本量规划研究

研究 3 主要探索不确定情境下的样本量规划, 可以为准确解决更实际、更广泛科研实践中的研究设计问题提供参考。主要思路是基于研究 1 选出的稳健性较好的拟合指标和研究 2 得到的基于检验力和效应量准确性分析规划样本量的范式, 拓展提出不确定性情境下样本量规划方法。然后, 基于该方法, 探索效应量不确定程度不同的条件下, 符合残差同质性的 MELSM 和完整 MELSM 的推荐样本量。

拓展提出的方法如下：

(1)定义效应量参数先验分布。根据已有研究得到效应量可能的取值范围, 再根据正态分布假设下, 可能的取值范围约为 6 倍标准差, 推算出效应量的标准差。由此, 推导出效应量先验分布的正态分布参数(也可以探索提出其他的先验分布)。

(2)抽取可能的效应量。从(1)中定义的先验分布里抽取 S 个效应量的值。

(3)生成样本。以所有备选模型中最复杂的模型为产生模型(Pek & Park, 2019), 以每个抽取的效应量的值为产生值, 其余参数设置参考研究 2, 生成 R 个样本量为 N 的样本, 共可得到 $R×S$ 个样本。

(4)建构模型。对 $R×S$ 个数据, 使用不同复杂程度的备选模型拟合数据, 应用研究 1 选出的较稳健的拟合指标选出合适的模型。基于每个效应量生成的数据集(共 R 个数据)计算检验力、效应量估计值的 95%可信区间宽度和 95%可信区间对真值覆盖率的平均值。

(5)整合结果。整合 S 个检验力和效应量准确性分析结果, 得到样本量为 N 时的检验力和效应量准确性指标的分布。

参考 Pek 和 Park (2019)的敏感性分析结果, 设定 S = 1000, R = 10000。将以上过程编写成 R 语言程序, 纳入本项目开发的软件包。

模拟研究沿用研究 2 的子研究分类, 分为 2 个子研究。

研究 3-1 探究不确定情况下符合残差同质性的 MELSM 的样本量规划。模拟条件包括均值模型自变量的效应量(0.2, 0.5, 0.8), 均值模型总体效应量不确定性程度(效应量分布的全距 = 0.15, 1.50, 3.00), 水平 1 样本量(10, 25, 50, 75, 100, 150, 200, 300)和水平 2 样本量(20, 30, 50, 75, 100, 150, 200, 300, 800), 共形成 3×3×8×9 = 648 种模拟实验条件水平的组合(即处理)。效应量分布的全距参考了 Pek 和 Park (2019)的模拟研究设置。参照基于

不确定性问题的检验力和效应量准确性分析基本方法，对于每种模拟实验条件水平的组合，基于模型 3 产生数据，通过比较备选模型 1~3 建构模型。最后，根据分析结果得到样本量为 N 时的检验力和效应量准确性指标的分布，研究者可以根据不同的标准确定推荐的样本量。可参考的标准如，检验力分布 20th 百分位数大于 0.8，或检验力均值大于 0.8 等。然后找到满足要求时对应的样本量，即为推荐值。

研究 3-2 探究不确定情况下完整 MELSM 的样本量规划研究。为简化研究，本研究固定均值模型中预测变量的效应量为中等效应，总体效应量不确定性为中等程度。模拟条件包括方差模型自变量的效应量(η_1 = 0.2, 0.5, 0.8)，方差模型总体效应量不确定性程度(效应量分布的全距 = 0.15, 1.50, 3.00)，水平 1 样本量(10, 25, 50, 75, 100, 150, 200, 300)和水平 2 样本量(20, 30, 50, 75, 100, 150, 200, 300, 800)，共形成 3×3×8×9 = 648 种模拟实验条件水平的组合(即处理)。参照基于不确定性问题的检验力和效应量准确性分析基本方法，对于每种模拟实验条件水平的组合，基于模型 6 产生数据。与研究 1 中建构完整 MELSM 的策略相同，先通过比较备选模型 1~3 确定均值模型，再通过比较备选模型 3~6 确定方差模型，实现模型建构。推荐样本量确定方式同研究 3-1。

最终，本研究将整合前三个研究的结果，开发简便易用的软件包，便于应用研究者实现 MELSM 的样本量规划、模型建构和数据分析等功能，促进新方法和新技术在心理学研究中的推广应用。

4.4　研究 4：混合效应均值−方差模型的应用研究

研究 4 将结合两个心理学的实际研究问题，演示 MELSM 的样本量规划、模型建构及结果解释的规范流程，验证前三个研究的结论在实际应用中的可操作性。两个案例的研究流程如下：(1)数据收集前，按照研究 3 确定的不确定性情况下样本量规划研究范式，基于前人相关研究结果，利用本项目所开发的软件包，输入多个备选模型、效应量的先验分布以及两个水平的样本量组合，从而根据软件包输出的各条件下效应量准确性及检验力指标结果，确定合理的水平 1 和水平 2 样本量。(2)按照所确定的样本量，完善研究设计，实施研究，收集数据。(3)基于数据建构合适的 MELSM，并估计参数，解释结果并作出结论。

案例 1 是基于 Stroop 研究范式的心理学实验研究，旨在探讨一致和非一致条件对正确作答反应时的影响，以及一致和非一致条件对被试正确作答反应时稳定性的影响。研究设计：实验内容为数字 Stroop 任务。包含一个被试内自变量，含有两个水平：一致条件和非一致条件。在一致条件中，字符的数量与显示的数字是匹配的(例如，333)。在非一致性条件中，字符的数量与显示的数字是不匹配的(例如，44)。实验任务是要求被试计算字符的数量，因变量为正确作答试次的反应时。收集的数据为嵌套结构，试次嵌套于被试。在该研究中，考虑的备选模型与模拟研究中的模型 1~模型 6 类似，均值模型和方差模型的自变量均为实验处理水平(一致/非一致)。

案例 2 是探索数学学习自我效能感和教师认知激发教学策略使用程度对数学成绩影响的教育心理学调查研究，旨在探讨学生数学学习自我效能感和数学教师认知激发教学策略使用程度对数学成绩的影响，以及二者对班级内成绩一致性的影响。研究设计：首先，编制有关学生数学自我效能感和数学教师认知激发教学策略使用程度的问卷，并通过预试验证问卷的信效度。其次，针对四川省某区县采取分层抽样的方法，首先抽取小学学校，然后在每所样本学校中随机抽取一个四年级班级的学生完成数学学习自我效能感问卷，其数学任课教师完成认知激发教学策略使用程度问卷，并统计这些学生在该地区统考中的数学成绩。收集的数据为学生嵌套于班级的结构。学生数学学习自我效能感为水平 1 自变量，数学教师认知激发教学策略使用程度为水平 2 自变量。在该研究中，考虑的备选模型包括(为简化研究暂不考虑自我效能感和认知激发使用程度的交互作用)：模型 1——均值模型自变量为自我效能感、认知激发使用程度，且不包含随机效应，符合方差同质性假设；模型 2——模型 1 基础上增加均值模型的随机截距；模型 3——模型 2 基础上增加均值模型中自我效能感的随机斜率；模型 4——模型 3 基础上增加方差模型的随机截距；模型 5——模型 4 基础上增加方差模型中自我效能感、认知激发使用程度的固定斜率；模型 6——模型 5 基础上增加方差模型中自我效能感的随机斜率。由于只有自我效能感是水平 1 自变量，因此仅该变量

可能有随机斜率。

5 理论建构与创新

适用于残差异质的情境，合理、深入挖掘嵌套数据信息的 MELSM 近年来受到国外研究者的高度关注。已有研究基于 MELSM 开展检验力分析等理论研究(Walters et al., 2018; Williams et al., 2019; Williams et al., 2021)，还有一些研究将 MELSM 应用于实证研究中，得到丰富的研究结果(Rast & Ferrer, 2018; Williams et al., 2020)。然而，目前国内外基于 MELSM 的理论研究和实践应用仍处于起步阶段，在实际应用中存在的样本量规划和模型建构等问题仍未得到充分解决，使研究者在应用MELSM时往往无所适从。为促进MELSM在心理学研究中的广泛应用，本研究将首先探究适用于完整 MELSM 模型选择的贝叶斯拟合指标，提出 MELSM 模型建构方法(研究 1)。然后探究 MELSM 在确定性情境下的基于检验力分析和效应量准确性分析的样本量规划方法(研究 2)，并进一步将研究 1 得到的较为稳健贝叶斯拟合指标应用于不确定情境下的样本量规划(研究 3)，最终形成 MELSM 的样本量规划和模型建构范式，并整合相关功能开发简便易用的软件包。最后，本研究拟通过实证研究验证模拟研究结果，演示 MELSM 的应用流程(研究 4)。

本研究拟提出的 MELSM 样本量规划和模型建构的理论范式如表 2 所示。在确定研究选题和完成实验设计之后，规范的实验研究通常包括样本量规划、数据收集、数据分析与结果解释等流程。一方面，研究者应当在数据收集前进行样本量规划，以确保样本量满足检验力和效应量准确性的要求。样本量规划存在三个方面的不确定性问题(Pek & Park, 2019)。基于蒙特卡洛的分析范式通过重复抽样能够处理样本变异造成的不确定性。当研究者能够基于预实验、前人研究或元分析等结果获得较有把握的效应量时，可不考虑总体效应量的不确定性；当研究者能够基于理论事先确定拟合模型时，可不考虑模型选择的不确定性。因此，根据研究是否存在这两种不确定性问题，研究者可采取不同范式开展样本量规划，确

表 2 MELSM 的样本量规划和模型建构的理论范式

步骤	数据收集前：样本量规划(研究 2，研究 3)			
	模型确定		模型不确定	
	效应量确定	效应量不确定	效应量确定	效应量不确定
第 1 步	根据先验信息确定 1 个效应量的值;	定义效应量参数先验分布，并从中抽取 S 个效应量的值;	根据先验信息确定 1 个效应量的值;	定义效应量参数先验分布，并从中抽取 S 个效应量的值;
第 2 步	基于待拟合模型生成 R 个样本量为 N 的样本，共可得到 R 个样本;	基于待拟合模型生成 R 个样本量为 N 的样本，共可得到 $R\times S$ 个样本;	基于备选模型中最复杂模型生成 R 个样本量为 N 的样本，共可得到 R 个样本;	基于备选模型中最复杂模型生成 R 个样本量为 N 的样本，共可得到 $R\times S$ 个样本;
第 3 步	基于待拟合模型拟合数据，计算检验力和效应量准确性;	基于待拟合模型拟合数据，计算检验力和效应量准确性;	基于各备选模型拟合数据，并根据贝叶斯拟合指标(研究 1)选择最佳模型的结果用于计算检验力和效应量准确性;	基于各备选模型拟合数据，并根据贝叶斯拟合指标(研究 1)选择最佳模型的结果用于计算检验力和效应量准确性;
第 4 步	整合结果，得到样本量为 N 时的检验力和效应量准确性指标的值。	整合结果，得到样本量为 N 时的检验力和效应量准确性指标的分布。	整合结果，得到样本量为 N 时的检验力和效应量准确性指标的值。	整合结果，得到样本量为 N 时的检验力和效应量准确性指标的分布。

步骤	数据收集后：模型建构(研究 1)	
	模型确定	模型不确定
第 1 步	直接拟合模型。	根据拟合指标确定最佳的均值模型;
第 2 步		根据拟合指标确定最佳的方差模型;
第 3 步		拟合模型选择得到的最佳 MELSM。

定后续数据收集时所需的样本量。另一方面, 数据收集完成后, 研究者应建构合适的模型。如果基于理论能够确定模型, 则无需进行模型选择, 可直接拟合模型并分析结果。如果存在模型选择的不确定性, 应当基于数据驱动的方法选择模型。具体而言, 研究者应根据本研究建议的贝叶斯拟合指标, 首先确定最佳的均值模型, 再确定最佳的方差模型, 从而得到最佳的 MELSM 用于数据分析。

本研究的创新之处主要体现在以下两个方面。

一是方法范式的创新。本研究充分考虑了适用 MELSM 的参数估计方法, 首次探讨贝叶斯估计框架下的拟合指标在 MELSM 模型选择中的表现及其影响因素, 并创新地提出在完整 MELSM 框架下依次对均值模型、方差模型进行选择的模型建构思路。这将为 MELSM 模型建构提供更可靠的拟合指标, 满足 MELSM 的应用需要。此外, 本研究将检验力分析和效应量准确性分析同时纳入样本量规划, 并在 Pek 和 Park (2019, 2022)的研究基础上, 改进不确定情境下的样本量规划范式, 总结出在效应量确定/不确定、模型确定/不确定 4 种情境下的样本量规划范式, 进一步完善 MELSM 的样本量规划方法, 丰富样本量规划的理论研究, 为实际应用中的样本量规划提供更加科学、可靠的方法学建议, 有助于提高实验研究的可重复性。

二是实践应用的创新。本研究结合心理学研究特点, 开发使用便利的软件包以满足贝叶斯框架下 MELSM 的样本量规划和模型建构的应用需要, 为 MELSM 的推广应用提供软件基础。这对科学开展研究设计、数据分析具有指导意义和创新价值, 是心理学实践领域创新性的探索。

综上, 本研究深入探索 MELSM 的模型建构和样本量规划方法, 为科学开展心理学研究提供方法支持。研究结果将进一步推动 MELSM 在心理学研究中的应用, 为深入挖掘嵌套数据的信息, 揭示复杂心理现象的本质提供全新视角。

参考文献

胡传鹏, 王非, 过继成思, 宋梦迪, 隋洁, 彭凯平. (2016). 心理学研究中的可重复性问题：从危机到契机. *心理科学进展*, *24*(9), 1504–1518.

温忠麟, 范息涛, 叶宝娟, 陈宇帅. (2016). 从效应量应有的性质看中介效应量的合理性. *心理学报*, *48*(4), 435–443.

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, *28*(11), 1547–1562.

Arend, M. G., & Schäfer, T. (2019). Statistical power in two-level models: A tutorial based on Monte Carlo simulation. *Psychological Methods*, *24*(1), 1–19.

Asparouhov, T., & Muthén, B. (2021). Advances in Bayesian model fit evaluation for structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(1), 1–14.

Baird, R., & Maxwell, S. E. (2016). Performance of time-varying predictors in multilevel models under an assumption of fixed or random effects. *Psychological Methods*, *21*(2), 175–188.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Blozis, S. A., McTernan, M., Harring, J. R., & Zheng, Q. (2020). Two-part mixed-effects location scale models. *Behavior Research Methods*, *52*(5), 1836–1847.

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, *23*(3), 389–411.

Brunton-Smith, I., Sturgis, P., & Leckie, G. (2017). Detecting and understanding interviewer effects on survey data by using a cross-classified mixed effects location-scale model. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *180*(2), 551–568.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), 9. https://doi.org/10.5334/joc.10

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Courvoisier, D., Walls, T. A., Cheval, B., & Hedeker, D. (2019). A mixed-effects location scale model for time-to-event data: A smoking behavior application. *Addictive Behaviors*, *94*, 42–49.

Depaoli, S., & Clifton, J. P. (2015). A Bayesian approach to multilevel structural equation modeling with continuous and dichotomous outcomes. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(3), 327–351.

Faul, F., Erdfelder, E., Lang, A. -G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program

for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Gelman, A., Meng, X. L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, *6*(4), 733–760.

González, B. J., de Boeck, P., & Tuerlinckx, F. (2014). Linear mixed modelling for data from a double mixed factorial design with covariates: A case-study on semantic categorization response times. *Journal of the Royal Statistical Society: Series C*, *63*(2), 289–302.

Halsey, L. G., Curran-Everett, D., Vowler, S. L., & Drummond, G. B. (2015). The fickle *p* value generates irreproducible results. *Nature Methods*, *12*(3), 179–185.

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics*, *64*(2), 627–634.

Hedeker, D., Mermelstein, R. J., Demirtas, H., & Berbaum, M. L. (2016). A mixed-effects location-scale model for ordinal questionnaire data. *Health Services and Outcomes Research Methodology*, *16*(3), 117–131.

Hoijtink, H., Mulder, J., van Lissa, C., & Gu, X. (2019). A tutorial on testing hypotheses using the Bayes factor. *Psychological Methods*, *24*(5), 539–556.

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications* (3rd ed.). New York, Routledge.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality & Social Psychology*, *103*(1), 54–69.

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: designs, analytic models, and statistical power. *Annual Review of Psychology*, *68*(1), 601–625.

Lee, W. Y. (2018). *Generalized linear mixed effect models with crossed random effects for experimental designs having non-repeated items: Model specification and selection* (Unpublished doctorial dissertation). Vanderbilt University.

Lin, X., Mermelstein, R. J., & Hedeker, D. (2018). A 3-level Bayesian mixed effects location scale model with an application to ecological momentary assessment data. *Statistics in Medicine*, *37*(13), 2108–2119.

Liu, X., & Wang, L. (2019). Sample size planning for detecting mediation effects: A power analysis procedure considering uncertainty in effect size estimates. *Multivariate Behavioral Research*, *54*(6), 822–839.

Martínez-Huertas, J. Á., Olmos, R., & Ferrer, E. (2021). Model selection and model averaging for mixed-effects models with crossed random effects for subjects and items. *Multivariate Behavioral Research*, *59*(4), 390–412.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.

Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. *Psychological Methods*, *9*(2), 147–163.

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Almenberg, A. D., ... Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, *73*, 719–748.

Park, J., & Pek, J. (2022). Conducting Bayesian-classical hybrid power analysis with R package hybridpower. *Multivariate Behavioral Research*. https://doi.org/10.1080/00273171.2022.2038056

Pek, J., & Park, J. (2019). Complexities in power analysis: Quantifying uncertainties with a Bayesian-classical hybrid approach. *Psychological Methods*, *24*(5), 590–605.

Rast, P., & Ferrer, E. (2018). A mixed-effects location scale model for dyadic interactions. *Multivariate Behavioral Research*, *53*(5), 756–775.

Schultzberg, M., & Muthén, B. (2018). Number of subjects and time points needed for multilevel time-series analysis: A simulation study of dynamic structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(4), 495–515.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B (Statistical Methodology), 64*(4), 583–639.

Thoemmes, F., MacKinnon, D. P., & Reiser, M. R. (2010). Power analysis for complex mediational designs using Monte Carlo methods. *Structural Equation Modeling*, *17*(3), 510−534.

van Erp, S., & Browne, W. J. (2021). Bayesian multilevel structural equation modeling: An investigation into robust prior distributions for the doubly latent categorical model. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(6), 875−893.

Walters, R. W., Hoffman, L., & Templin, J. (2018). The power to detect and predict individual differences in intra-individual variability using the mixed-effects location-scale model. *Multivariate Behavioral Research*, *53*(3), 360–374.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on *p*-values: Context, process, and purpose. *The American Statistician*, *70*(2), 129–133.

Williams, D. R., Martin, S. R., Liu, S., & Rast, P. (2020).

Bayesian multivariate mixed-effects location scale modeling of longitudinal relations among affective traits, states, and physical activity. *European Journal of Psychological Assessment*, *36*(6), 981–997.

Williams, D. R., Martin, S. R., & Rast, P. (2022). Putting the individual into reliability: Bayesian testing of homogeneous within-person variance in hierarchical models. *Behavior Research Methods*, *54*(3), 1272–1290.

Williams, D. R., Mulder, J., Rouder, J. N., & Rast, P. (2021). Beneath the surface: Unearthing within-person variability and mean relations with Bayesian mixed models. *Psychological Methods*, *26*(1), 74–89.

Williams, D. R., Zimprich, D. R., & Rast, P. (2019). A Bayesian nonlinear mixed-effects location scale model for learning. *Behavior Research Methods*, *51*(5), 1968–1986.

Zhang, Z. (2014). Monte Carlo based statistical power analysis for mediation models: Methods and software. *Behavior Research Methods*, *46*(4), 1184−1198.

Model construction and sample size planning for mixed-effects location-scale models

LIU Yue[1], FANG Fan[1], LIU Hongyun[2,3], LEI Yi[1]

([1] *Institute of Brain and Psychological Sciences, Sichuan Normal University, Chengdu* 610066, *China*)

([2] *Beijing Key Laboratory of Applied Experimental Psychology, Beijing Normal University, Beijing* 100875, *China)*

([3] *Faculty of Psychology, Beijing Normal University, Beijing* 100875, *China*)

Abstract: With the development of data-collection technics and increasing complexity of study designs, interest in Mixed-Effects Location-Scale Models (MELSM) has increased drastically. When residual variances are heterogeneous, these models are able to add predictors in different levels, then help explore the relationship among traits and simultaneously investigate the inter- and intra-individual variability, as well as their explanatory variables. This project includes both simulated studies and empirical studies. In detail, the main contents of this project are: 1) Comparing and selecting candidate models based on Bayesian fit indices to construct MELSM; 2) Planning sample size according to both power analysis and accuracy in parameter estimation analysis for MELSM; 3) Extending the sample size planning method for MELSM to better frame the considerations of uncertainty; 4) Developing an R package for MELSM and illustrating the application of MELSM in empirical psychological studies. Based on the project, we hope these statistical models can be widely implemented. Moreover, the reproducibility and replicability of psychological studies will be enhanced finally.

Keywords: nested data, mixed-effects location-scale models, model construction, sample size planning